

A STUDY OF THE VOCABULARY OF RADIO

BY

DONALD E. HARGIS

Reprinted from
Vol. XII, Research Annual, 1945
SPEECH MONOGRAPHS



A STUDY OF THE VOCABULARY OF RADIO*

DONALD E. HARGIS

University of Michigan

THE PROBLEM

FOR centuries scholars and investigators have examined the basic unit of speech and writing: the word. Word studies based on definition, based on finding the vocabulary; those designed to find levels, and those plotting vocabulary growth are numberless. In the multitude of these studies one type, the objective word count, frequently is to be found. The objective or quantitative word count is a tabulation of the words in selected written materials and a listing of those words in the order of frequency of occurrence.

The early attempts with word lists were to supply English speaking persons with a "beginner's vocabulary" in a foreign language and to establish systems of shorthand. In the years following these first studies hundreds of word counts have been made, some of them general and some highly specialized. With the growing interest in word counts as an aid to teaching foreign languages, the American and Canadian Committees on Modern Languages took a hand in the promotion of research.

In the English language the word count method has been extensively used as the basis for devising spelling lists. The early investigations of reading vocabulary were all by the empirical method, and it was not until after 1910 that objective studies were attempted. Perhaps the best known and the most widely used of the word lists growing out of the frequency word count studies

is that of Thorndike.¹ Up to the present time it remains as the outstanding, single, original work in the field. "Thorndike's list has been almost universally considered as the ultimate in the matter of word counting. Nearly every subsequent study has used it as a basis for comparison and as a model for the technique of counting."²

His investigation was originally published as a list of 10,000 different words³ which were found to occur most frequently in 4,565,000 running words taken from forty-one sources. The list was constructed to aid in the teaching of reading in the elementary schools. Since the publication of the original study, Thorndike has made a count of 5,000,000 more running words from over two hundred other sources and has included material from other studies.⁴ From this last material in addition to the original study, he revised the list and extended it to 20,000 different words.

Only two studies have been made in the field of the vocabulary of radio. The first is by Dale⁵ on the vocabulary level in a particular field of radio speaking. He was attempting to find the vocabulary level of radio speeches aimed at various grades in school.

The second study, by Smith,⁶ is an

¹ Thorndike, E. L., *Teachers' Word Book of 20,000 Words* (New York, 1931).

² Fries, C. C., and Traver, A. A., *English Word Lists* (Washington, D. C., 1940), p. 23.

³ Thorndike, E. L., *Teachers' Word Book* (New York, 1921).

⁴ Thorndike, E. L., *A Teachers' Word Book of 20,000 Words* (New York, 1931).

⁵ Dale, E., "The Vocabulary Level of Radio Addresses," *Education on the Air* (Columbus, Ohio, 1931), p. 245.

⁶ Smith, S. S., "Radio Vocabulary," *QJS*, XXVIII (1942), p. 1.

*A condensation of a dissertation submitted in partial fulfillment of the degree of Doctor of Philosophy at the University of Michigan.

attempt to discover the vocabulary of radio and, specifically, the level of the vocabulary used. The study seems to have been made on the general frequency count method with the special listing of words above certain levels in the Thorndike Word List.

One of the basic faults of the work in the past is that it has been done, almost without exception, on material that was written to be read silently. The resulting word lists are not representative of the vocabulary in everyday use, but of the "literary" vocabulary of authors whose ideas have been put into print. Doubtless, for the present at least, there have been enough "word counts" in the field of written material.

The first necessity in a new study is to limit the source material so that the final word list will be useful. In radio material, which stands somewhere between the printed, written-to-be-read-silently, and conversational speech which is not written at all, we have a useful source for a word count. The following two factors recommend radio scripts as a source for a quantitative word count: the material is close to the spoken word and it is current in conception. Radio scripts further recommend themselves in that no studies of any scope have been made on this type of material. For that reason the present study is not a mere repetition of what has been done before.

An objective analysis is needed in radio for several reasons. First, it could serve as a guide for writers in the field. From the word list, writers could see which words are used and which shunned by other writers. It would serve to objectify the approach to writing. Second, this study would have value in the teaching field. When one makes a survey of the text-books in the field of radio

writing, he finds references in all of them to the vocabulary. These observations range from simple warnings to use a non-complex vocabulary to lengthy discussions of the type of vocabulary which should be used. Writers, such as Abbot,⁷ Barnouw,⁸ Dale,⁹ Carlile,¹⁰ Whipple,¹¹ and Hettinger,¹² to mention only a few, constantly reiterate warnings that the radio vocabulary must be simple, non-complex, and gauged to the level and diversity of the radio audience.

It is easy to see the concern with which the teachers and the text-book writers view the problem of the vocabulary in radio writing. The point which is obvious in their observations is that their knowledge of that vocabulary has been gathered from practical writing experience, from radio production experience, or from observations as listeners. As valuable as the compilation of the results of such experience may be, their observations can lead only to general statements because of their subjective approach to the problem. With a word list made up by an objective count of radio materials, the writer and teacher could have a scale on which to base the discussion of vocabulary use.

Finally, the study should have a place in the analysis of the effect of the radio on the general public. A list, such as the one proposed, would provide an objective guide by which one could discover what is present on the air and, by comparing the list with other lists,

⁷ Abbot, W., *Handbook of Broadcasting*, revised (New York, 1941), p. 86.

⁸ Barnouw, E., *Handbook of Radio Writing* (Boston, 1939), p. 66.

⁹ Dale, *op. cit.*, p. 247.

¹⁰ Carlile, J. S., *Production and Direction of Radio Programs* (New York, 1939), p. 260.

¹¹ Whipple, J., *How to Write for Radio* (New York, 1938), p. 24.

¹² Hettinger, H. S. and Neff, W. J., *Practical Radio Advertising* (New York, 1938), p. 189.

form some idea of that vocabulary level. At the same time, comparisons could be made between types of programs. Differences in vocabulary of serials, newscasts, comedy-variety programs, and talks could be uncovered and analyzed. These differences, if existing, could be correlated with the general radio vocabulary list or outside lists. In that way some definite and objective judgment could be passed on the materials on the air and the probable influence which they would have on the speech habits of the listener.

The problem, then, was the construction of a quantitative word list compiled from selected radio materials which would be representative of the general vocabulary used in radio writing. This list was to be a count of all of the words which occur in such materials, arranged according to frequency and range of appearance. The word list so collected was to be compared and analyzed on the basis of other existing lists. In that way, the comparative level of the radio vocabulary was to be estimated; and, in addition to this, comparisons were to be made between the types of radio material used in the study.

MATERIALS AND METHODS

In the studies which have been made a wide divergence of method is to be found within the framework of the frequency count. The problem was to determine what methods and materials were to be used in the present study. The fundamental problem centered about the selection of materials. The representative nature of the materials, the inclusion of all different types of scripts, and the universality of those materials, the size of the audience and the number of stations on which they were broadcast had to be considered; materials most universal in terms of

the listener should be used. Another factor was the timeliness of the material; the current value of the radio material is important if a true picture is desired. From the point of view of research in the social sciences, there were other considerations as to sampling besides universality and representativeness. Such matters as the time of the year, the time of the day, the authors of the materials, the sections of the country represented, the station or network selected, and the statistical correlatives of these problems were important. To work out all of the aspects of the sampling problem in the selection of radio materials for such a study as this would have been, however, a separate investigation in itself.

On the basis of universality, network broadcasts were selected. It was decided that for practical purposes programs broadcast on one network of the four we have in this country differ very little from those broadcast on another. Hence, merely by chance, the National Broadcasting Company was selected. Through the courtesy of the manager of the script division, arrangements were made to secure the necessary scripts. It was determined that all of the broadcasts made in single days would be used. On a purely random basis three days in December, 1942—a Tuesday, a Thursday, and a Sunday—were chosen. Over one hundred different scripts were included in the three days of broadcasting, and within this material were all of the different types of programs which go on the air. The material lived up to the criteria set down for universality, proper proportion of types, and practicability, and thus was assumed to be an acceptable sample.

The basic method of the study was a simple compilation and listing of all of

the words found in all of the scripts. All nouns were separated into singular and plural numbers, all verbs entered in the tense used, and all other variations of words accounted for. In addition to this, homographs, being in reality different words, were treated as such. All slang, contractions, abbreviations, and colloquial speech were listed. This gave greater accuracy to the study than would have been the case if the forms had not been listed separately. However, as this was not a semantic study, no effort was made to separate the different meanings of a word.

A frequency figure, the number of times the word occurs, was tabulated for each different word in each script, and a total frequency figure determined from that data. Also a range figure, the number of different scripts in which the word appeared, was calculated. As all other important frequency word studies, such as Horn¹³ and Thorndike,¹⁴ recognized both frequency and range in compiling the word list, both factors were examined here. However, it was not assumed at the outset that they were independent factors, as all studies except that of Vander Beke¹⁵ had done. A correlation was run between frequency and range; if they were found to be independent, both would then be used in compiling the final list. All of the different words were listed in alphabetical order with both the frequency and the range figure indicated.

On the basis of the list several operations were completed to find the meaning of the material. First, an analysis was made of the words common to all

of the materials examined. At the same time some indication of the nature of the most frequent words was considered valuable; such things as grammatical nature and function, etc. were analyzed. From this the study proceeded to an analysis of the types of words found on the various levels. This study was based on grammatical usage; thus, the most common parts of speech were discovered. Another check was on the differences between types of programs. It was comparatively simple to classify the programs into the general level of the vocabulary of each group. The significance of the percentages of common words could be calculated and further check made on the reality of any differences between the various types of programs. As the Thorndike list is the best existing and most commonly used yard-stick for word studies, it was used here as a basis of comparison. Whatever differences or likenesses were found between the present list and Thorndike's could be interpreted as differences in the materials used in the two studies. Something of the comparative level of the two lists could, thus, be shown.

Final check was then made on the reliability of the list. These calculations were made by splitting the data on the words into halves, thirds, and quarters, and by running correlations. Finally, these results were worked together into a summary and conclusions drawn from them, as well as suggestions made for further study.

THE WORD LIST

In the materials examined 227,269 running words were present with 5,719 different words. The range varied from 208 to 1. The correlation between range and frequency was found to be $\pm .89 \pm .0019$. This correlation was high enough to demonstrate that the factors

¹³ Horn, E., *A Basic Writing Vocabulary* (University of Iowa, 1926).

¹⁴ Thorndike, E. L., *A Teachers' Word Book of 20,000 Words* (New York, 1931).

¹⁵ Vander Beke, G. E., *French Word Book* (University of Wisconsin, 1924).

of range and frequency are not independent but operate in the same way. Hence, the frequency of the word was used to determine its place and rank in the listing. It was found that the first four thousand different words of greatest frequency divided into thousand groups. At the same time, the thousand words of greatest frequency divided naturally into groups of one hundred words each. The thousand of second greatest frequency could not be broken down beyond two groups of five hundred words each. The third and fourth thousands could not be divided, nor could the last group of 1,719 words of the least frequency. The frequency varied from one word with a frequency of 7,561 down to 611 words with a frequency of one. In listing the words, first, the grammatical usage of each word was indicated by symbol; second, the frequency was given; third, the range; fourth, the index number for the word showing its placement into hundred and thousand word groups in this list; and last was given the index number for the word in the Thorndike list.

RESULTS AND ANALYSIS

In the 129 scripts and their commercials which were counted there were 227,269 running words with 5,719 different words. Dividing the number of running words by the number of different words, a ratio of 39.7 was obtained; this means that there is one new word in every 39.7 running words. The average ratio for nine other representative studies done on a similar basis is 49.2. The ratio found in the present study is close to the mean for all of the studies reviewed, varying from it by only 9.5 points.

The programs included in the count were classified as to type into ten divisions: serial drama, music programs,

dramatic programs, educational-talk programs, sports programs, comedy-variety programs, news broadcasts, religious programs, commercial announcements, and a miscellaneous group which could not be grouped with any of the others and which were single examples of their types. The data for the types are given in Table I.

In the percentage of programs the serial drama, music, and news program made up seventy-five per cent of the number of different scripts. When percentages of running words were compared, over sixty-five per cent of the words came from serial drama, comedy-variety, news programs, and commercial announcements. As the program of greatest length, measured in running words, should be the most important in its effect on the listening public, this second method of comparison is the more valid. The serial drama becomes the most important type along with the commercial announcement, which is close to it in percentage of running words. Difficulty is encountered in trying to analyze the importance of the music program by the number of running words. One cannot measure music in terms of words, and so this criterion of measurement fails. It is interesting to note that using any of the criteria the serial drama stands in first place. It seems safe to assume that it holds the most important place in the sample of materials. By using all of the data available it would seem that serial drama should rank first in importance; music should rank second; news and comedy-variety programs are third; and the other types relatively unimportant as they total less than twenty-five per cent of the whole on any basis of comparison. It is relatively impossible to gauge the position of the commercial an-

TABLE I
TYPES OF PROGRAMS

Type	Number	Percentage of Total Number	Time on the Air in Minutes	Percentage of Total Time on the Air	Number of Running Words	Percentage of Running Words	Number of Different Words	Percentage of Different Words	Ratio of Running to Different Words
Serial drama	47	36.4%	720	28.8%	44,205	19.4%	834	14.5%	53.0
Music	26	20.1%	555	22.2%	19,164	8.4%	923	16.1%	20.7
Drama	4	3.1%	120	4.8%	13,101	5.7%	832	14.5%	15.7
Ed.—talk	8	6.2%	165	6.6%	18,410	8.1%	1107	19.3%	16.6
Sports	3	2.3%	45	1.8%	6,013	2.6%	660	11.5%	9.1
Comedy-variety	10	7.7%	330	13.2%	28,613	12.5%	518	9.0%	55.2
News	23	17.8%	320	12.8%	37,505	16.5%	1582	27.6%	23.7
Religious	2	1.5%	60	2.4%	6,767	2.9%	239	4.1%	28.2
Commercial	79	—	—	—	41,491	18.2%	622	10.8%	66.7
Unclassified	6	4.6%	180	7.2%	12,000	5.2%	677	11.8%	17.7
Totals	208	100%	2495	100%	227,269	100%	—	—	—

nouncement, except to say, that on the basis of the number of words used, it ranks at the top and coequal to serial drama.

The news broadcast includes more different words than any other type. It includes over twenty-five per cent of all of the different words found in the study, and has 8.3% more of them than any other group of scripts. The educational-talk program is second with 19.3% of the different words, and the music broadcast third with 16.1%. The three types mentioned above have the largest vocabularies in comparison with the other groups. Serial drama and drama have the same percentage of different words, 14.5%, while the comedy-variety program is the lowest with 9.0%.

The serial drama and comedy-variety programs and commercial announcements have the high ratios of 53.0, 55.2, and 66.7 of running to different words respectively. Music and drama have ratios of 20.7 and 15.7; news and religious programs are slightly higher at 23.7 and 28.2. The lower this ratio the more different words there are in com-

parison to running words, and, hence, the larger the vocabulary in a given number of running words. The vocabularies of serial drama, comedy-variety programs, and commercial announcements, as expressed by the ratio, are low, and those of the others, larger.

Of the first one hundred words in the present list, all but eighteen appear in the Thorndike list of 135 common words. Two of these eighteen are not infinitives of the verb, eight are contractions, three are paired words which appear twice in different grammatical forms in the list. So, in reality, there are seventy-nine words which are common to both lists. All but eight of the one hundred most frequent words in the present list are in the five hundred most frequent words of the Thorndike list. Of these eight, seven are contractions and the eighth a noun. Considering that there appears to be some confusion as to how Thorndike counted contractions, the variations are slight from list to list. Only fourteen words in the present list have a range figure of 208 which indicates that they were present in every script and every commercial. Those

words are: "the," "to," "and," "a," "you," "of," "in," "for," "that," "is," "that," "this," "with," and "are." "That" is listed both as an adjective and as a pronoun and so appears twice. Two of these words are articles; five are prepositions; one, a conjunction; three, pronouns; one, an adjective; and two, verbs. This list of common words corresponds very closely with the common words noted in other similar studies.

An analysis of the parts of speech was undertaken as the next step. Nouns and verbs have a prominent place in the number of different words; over seventy-five per cent of the different words are nouns and verbs. Those two groups along with adjectives account for ninety per cent of the different words. In number and percentage of running words the story is slightly different. Nouns and verbs, especially nouns, lose some of the importance which the first basis of comparison gave to them. Adjectives drop slightly in importance in relation to other parts of speech. The most striking change is the increase in the importance of pronouns. They include only 1.2% of the different words, while they comprise 13.0% of the running words. In the same way, prepositions, conjunctions, articles, adverbs, and contractions gain in importance. The noun and the verb are still the most frequent by almost twice the total of any other part of speech. The change in relative importance of all of the parts of speech is the interesting one.

Articles, of which there are only three in the list, make up 5.7% of all of the running words. On the other hand, nouns, which make up 46.4% (2,656, of the different words) represent only 21.9% of the running words. It is this frequent use of a few words which is important. Nouns and verbs, which are

over seventy-five per cent of the different words, include only forty-four per cent of the running words. Contrary-wise, prepositions, conjunctions, articles, contractions, interjections, and pronouns total but four per cent of the different words, and yet include thirty-seven per cent of the running words. One can see from this how the frequent repetition of a few words accounts for a large part of the running words. That these words should be the parts of speech which they are, the common connectives, etc., is also noteworthy: it demonstrates that a large part of the writing is built of words which do not express the meanings and ideas, but which help along the meanings expressed by the verbs, nouns, and adjectives. As over sixty-five per cent of the running words are nouns, verbs, adjectives, and pronouns, one can say that they appear to be the most important groups of words in terms of "total" usage.

An analysis of the parts of speech leads to the general conclusion that the most frequent parts of speech, both in different and running words, are the conjunctions, articles, prepositions, and pronouns. That is, those parts of speech have a higher percentage of both different and running words in the thousand-most-frequent words in the distribution scale, than do the nouns, verbs, and adjectives. The distribution of the adverbs, interjections, and contractions in the thousand words of greatest frequency does not allow for any general conclusion. It is evident that a relatively few different conjunctions, articles, prepositions, and pronouns make up the bulk of the most frequent words. This bears out what other researchers in the field have found.

In the study of the detailed distribution of the words by program types, the

first observation to be made is the high percentage of running words in each type which come from the one hundred words of greatest frequency. Except in the case of the sports program, over thirty per cent of the running words in each type comes from this group. The comparison of the percentages of different words throughout the distribution does not reveal the same facts, however. At the upper end of the distribution, in the first thousand different words, the serial drama, and comedy-variety shows, and the commercial announcement have higher percentages of different words than do the music, drama, educational-talk, sports, religious, and news broadcasts. From the group of 1,719 words of least frequency, the music, drama, educational-talks, and news broadcasts have a relatively high percentage of different words. This difference in different words at the two ends of the distribution scale for the various types of programs indicates that the serial drama, comedy-variety programs, and commercial announcements use more words which are repeated many times than do the music, drama, educational-talk, and news broadcasts. This confirms the conclusion that the latter four types of programs have larger vocabularies than the former three.

In an analysis of the parts of speech in relation to the program types, the following general conclusions can be drawn. Nouns are the most common words in all types of programs on the basis of different words, except for the comedy-variety broadcast, where adjectives are more prominent. In the per cent of running words the noun is the most common in the serial drama, comedy-variety programs, and commercial announcements. For the music, educational-talks, sports, and news programs the

percentage of verbs to running words is the largest. The percentage of pronouns to running words is highest in the drama and religious broadcasts. All types are comparatively low in percentage of different prepositions, conjunctions, articles, adverbs, interjections, and contractions. The different adjectives play a considerable part among the different words used in all types. The bulk of the running words in all types are nouns, verbs, adjectives, pronouns, and prepositions in approximately that order of importance. There are individual variations within the types, but this is the general trend.

The words in the list were reclassified and arranged on the basis on which Thorndike¹⁶ made up his list. All forms of the verb were placed under the infinitive, all forms of the noun under the singular, etc. This new list included 4,235 different words. Of these words, three hundred were not to be found in any form in the Thorndike list and were set aside. The list which was compared with the Thorndike scale had 3,935 different words. These words were regrouped into frequency order and divided into groups of five hundred words as Thorndike did with the first 5,000 words in his list. A correlation of $+0.69 \pm 0.0056$ was found between the Thorndike list and the present revised one. This figure is high enough to indicate a substantial relationship between the placement of the words in the two lists. It demonstrates that in the samples used in the two pieces of research the frequency of the words, which in general determined the place order in both lists, was relatively the same. The most frequent words were most frequent in both lists,

¹⁶ See Thorndike, *A Teachers' Word Book of 20,000 Words*, pp. iv and v for Thorndike's method of compiling the list.

TABLE II
RELIABILITY OF MATERIAL

Correlation between:	Reliability of third:	Reliability of whole:
First and second third	$+.74 \pm .009$	$+.89 \pm .007$
First and third third	$+.61 \pm .013$	$+.81 \pm .012$
Second and third third	$+.47 \pm .016$	$+.72 \pm .026$

TABLE III
RELIABILITY OF MATERIAL

Correlation between:	Reliability of third:	Reliability of whole:
December 3 and December 6	$+.61 \pm .013$	$+.81 \pm .012$
December 1 and December 3	$+.76 \pm .009$	$+.90 \pm .005$
December 1 and December 6	$+.51 \pm .015$	$+.78 \pm .02$

and the least frequent words were least frequent in both. While the relationship is by no means a perfect one, it is high enough in relation to the size of the sample used in the present study to be indicative of more than a casual trend.

Sixty-nine and six-tenths per cent of all the words in the present study found in the Thorndike list occur in his 5,000 most frequent words; 88.7% appear in the 10,000 most frequent; and 96.5% in the 15,000 most frequent words. This lends significance to the conclusion that the vocabulary as expressed in the present list is low in terms of the Thorndike scale. The three hundred words in the present list which were not in Thorndike's are mainly words of high current usage, slang, and contractions; this would account for their absence in his list.

As a check on the reliability of the materials in the study, correlations were run on the frequencies of split halves and thirds of the 1,000 words of greatest frequency. The words were split arbitrarily on the basis of programs. Two new figures of frequency were worked out for each word in the thousand of greatest frequency. The correlation between these two halves was $+.59 \pm .013$.

By the Spearman-Brown formula the coefficient of reliability for all of the material was $+.74 \pm .017$, which is marked enough to indicate a high reliability for the material. In the same way the material was split into thirds and correlations run between the 1,000 words of greatest frequency. The results of these correlations are shown in Table II. The same type of correlation was made between the 1,000 words of greatest frequency in the three different days from which the sample of material was taken. These correlations are shown in Table III. These correlations show that there is a high degree of reliability in the material as measured by this method.

Using the Spearman-Brown formula¹⁷ for the effect upon reliability of increasing the size of the sample, the effect of doubling or tripling the amount of material in the sample was determined, based on the correlations made above. If the formula is applied to the figure for the split half correlation, the reliability of double the amount of material in the present sample would be $+.84 \pm .006$. By the same method, the effect of tripling the amount of material

¹⁷ Garrett, H. E., *Statistics in Psychology and Education* (New York, 1941), pp. 315-316.

would be to increase the reliability to $+ .89 \pm .004$. The formula was also used to determine the increase in the amount of material necessary to give a certain predetermined coefficient of reliability. If a coefficient of reliability of $+ .90$ is desired, then there must be three times as much material as at present. If the coefficient is to be increased to $+ .95$, then six times as much material must be used. These demonstrate the high degree of reliability for the sample used in the present study and show the amounts of material which will have to be used in future studies if higher reliability coefficients are desired.

CONCLUSIONS

1. In the present study the method used for selecting materials and the materials themselves appear to be highly satisfactory. A careful examination of the materials included in the sample reveals that they are surprisingly representative of what goes on the air.

2. The correlation of frequency and range to determine the *modus operandi* for assembling the word list demonstrates that both range and frequency need not be taken into account in arranging the words in the list if the materials are homogeneous. In the same way, the other major departure in method, the inclusion of every form of every word exactly as it appeared in the materials, seems to be vindicated. The different placements and frequencies for homographs in the present list justifies the use of the plan.

3. The various types of programs do differ in vocabulary. (a) The serial drama and comedy-variety programs and the commercial announcements have smaller vocabularies as expressed in the ratio of running to different words than do the music, drama, educational-talks, and news broadcasts. (b) Music,

drama, educational-talks, and news broadcasts have a vocabulary which differs from the other types in that they include more different words which are used but a relatively few times each. It is safe to assume that the more different words of low frequency used and the more different words used, the larger and more complex the vocabulary level of the material. Thus, the music, drama, educational-talks, and news programs have larger and more complex vocabularies than the serial drama, comedy-variety programs, and commercial announcements. (c) The music, drama, educational-talks, and news broadcasts use higher proportions of the meaningful words—nouns, verbs, and adjectives—than do the other types of programs; the serial drama, comedy-variety programs, and commercial announcements use fewer different words and repeat those words over and over many times. This latter group uses higher relative proportions of the less important connectives and the like than does the former group. This would tend to demonstrate that the meanings and exact ideas were more carefully expressed in the music, drama, educational-talks, and news broadcasts.

4. The words of greatest frequency are conjunctions, prepositions, articles, and adverbs. This confirms the findings of other researchers who have discovered that the function words are the most common ones in terms of frequency of occurrence. There are comparatively few different nouns, verbs, and adjectives among the high frequency items in the list. High percentages of different nouns, verbs, and adjectives appear in the 1,719 words of least frequency. It is impossible from the sample to determine the true relative place of the words in this group. They ap-

pear so infrequently that one cannot be sure whether they are important or not. As they are the most important items for meaning, and as the high frequency of the function words seems quite well established, research on a wide range of materials should be made on the nouns, adjectives, and verbs in order to establish more definitely their true frequency positions.

5. The comparisons of the present list with the Thorndike list provide a satisfactory method of establishing levels. Two conclusions are available from the comparisons made between them. (a) The correlation between the placement of the words in the two lists indicates that there is a high degree of agreement in the frequency placement of the different words in the two. There is fairly close agreement as to the different words used. However, there are three hundred words in the present study which did not occur in Thorndike's list. Since this is only 7.0% of the words in this list as revised to fit the Thorndike list, the correspondence between the two, both on this basis and the correlation, is high enough so that it can be said that there is a substantial relationship between them. (b) The placement of the words in the present study in the distribution of the Thorndike list suggests rather positively that the vocabulary of the sample used in this study is low in comparison to the vocabulary of the sample used by Thorndike. There is a high

representation of the words from the most frequent thousand in the Thorndike list and a very marked and steady decrease in the number of words from each of the next thousand of lesser frequency up through the 20,000 of least frequency. This suggests that if the radio writer is to use the Thorndike list as a **check-list for writing**, he should beware of words above the 10,000 word level, as only twelve per cent of the words in the present list are from above that level.

6. The correlations to establish reliability demonstrate a high degree of homogeneity for the sample of radio material. The application of the prediction formula shows that if the amount of material were doubled or tripled a very high coefficient of reliability would be attained. At the same time, there should be quite high correlation between the present sample and any other general radio sample chosen.

SUGGESTIONS FOR FURTHER STUDY

1. Further study of the same type should be carried on using more varied radio sources.
2. Further detailed comparison should be made with Thorndike's list.
3. Profitable research could be made in the grammatical usage of radio writing.
4. A study could be made of radio materials exactly as they are delivered on the air.

